Implications of Game Triads for Observations of Trust and Reciprocity

By James C. Cox

University of Arizona
jcox@bpa.arizona.edu

First draft: September 1999. This draft: December 2000

Abstract

This paper develops a triadic design for conducting trust and reciprocity experiments. A large literature on single-game trust and reciprocity experiments is based on the assumption that subjects' utility payoffs are the same as their own monetary payoffs in the experiments. Such designs test compound hypotheses that include the hypothesis that other-regarding preferences do not affect behavior. In contrast, experiments with the triadic design do discriminate between transfers resulting from trust or reciprocity and transfers resulting from other-regarding preferences. Decomposing trust from altruism and reciprocity from altruism is critical to obtaining empirical information that can guide the process of constructing models that can increase the empirical validity of game theory.

Keywords: experimental economics, game theory, trust, reciprocity, altruism

JEL Classification: C70, C91, D63, D64

Implications of Game Triads for Observations of Trust and Reciprocity*

1. Introduction

In their seminal work on game theory, von Neumann and Morgenstern (1944, 1947) thought it necessary to simultaneously develop a theory of utility and a theory of play for strategic games. In contrast, much subsequent development of game theory has focused on analyzing the play of games to the exclusion of utility theory. In the absence of a focus by game theorists on utility theory, it is understandable that experimentalists testing the theory's predictions have typically assumed that agents' utilities are affine transformations of (only) their own monetary payoffs in the games. This interpretation of game theory incorporates the assumptions that agents do not care about others' (relative or absolute) material payoffs nor about their intentions. There is a large experimental literature based on this special-case interpretation of the theory, which I shall subsequently refer to as the model of "self-regarding preferences." The part of the literature concerned with public goods experiments and trust and reciprocity experiments has produced replicable patterns of inconsistency with predictions of the model of self-regarding preferences. For example, the patterns of behavior that have been observed in one-shot trust and reciprocity games are inconsistent with the subgame perfect equilibria of that model. But this does *not* imply that the observed behavior is inconsistent with game theory, which is a point that has not generally been recognized in the literature.

In one prominent research program, the central empirical question has been posed as a contest between game theory and alternative theories based on ideas of cultural or biological evolution.[1] For example, McCabe, Rassenti, and Smith (1998) pose the question as follows:

> *Our objective is to examine game theoretic hypotheses of decision making based on dominance and backward induction in comparison with the culturally or biologically derived hypothesis that reciprocity supports more cooperation than predicted by game theory (p. 10)…*

and state their conclusion as

Contrary to noncooperative game theory, but consistent with the reciprocity hypothesis, many subjects achieve the symmetric joint maximum under the single play anonymous interaction conditions that are expected to give game theory its best shot (p. 22).

Another distinguished research program has focused on inconsistencies between the predictions of principal-agent theory and behavior in experimental labor markets.[2] For example, Fehr, Gächter, and Kirchsteiger (1997, p. 856) conclude that

Our results indicate, however, that the neglect of reciprocity may render principal agent models seriously incomplete. As a consequence it may limit their predictive power. Moreover, the normative conclusions that follow from models that neglect reciprocity may not be correct.

Widely-disseminated conclusions about robust observations of trust and reciprocity have motivated developments of utility theory intended to improve the empirical validity of game theory. For example, Rabin (1993) and Dufwenberg and Kirchsteiger (1998) have developed models that incorporate perceptions of others' intentions into the utilities of game players. In contrast, Levine (1998), Fehr and Schmidt (1999), and Bolton and Ockenfels (2000) have developed models that incorporate other-regarding preferences (or fairness) into game players' utilities. A model that incorporates both intentions and fairness has been developed by Falk and Fischbacher (1998). But there is a problem with the widely-disseminated conclusions about behavior that are motivating these theory developments: the conclusions are not all supported by the experimental designs that generated the data.

The present paper reexamines some central questions in the literature on trust and reciprocity. It specifically questions the widely-accepted conclusion stated in a recent survey article by Fehr and Gächter (2000b, p. 162):

> *Positive reciprocity has been documented in many trust or gift exchange games (for example, Fehr, Kirchsteiger and Riedl, 1993; Berg, Dickhaut, and McCabe, 1995; McCabe, Rassenti and Smith, 1996).*

The conclusion that positive reciprocity is "documented" by data showing that many proposers send, and responders give back money in trust and gift exchange games is not supported by the experimental designs in the cited papers. The source of the difficulty is that the single-game experimental designs used to generate the data in these experiments do *not* discriminate between actions motivated by reciprocity and actions motivated by altruism. In the present paper, a triadic experimental design is introduced that makes it possible to discriminate between transfers resulting from trust or reciprocity and transfers resulting from other-regarding preferences. Being able to discriminate between the implications of other-regarding preferences and trust or reciprocity is important to obtaining the empirical information that can guide the process of formulating a theory of utility that can increase the empirical validity of game theory.

2. Definitions

Interpretations of data in this paper will be based on the following definitions. Preferences over the absolute and relative amount of another individual's money payoff, in addition to one's own money payoff, will be referred to as "other-regarding preferences."[3] Such preferences can be altruistic or malevolent. They involve ideas of the fairness of outcomes. Let y^j denote the amount of income of agent *j*. Assume that agent *k*'s preferences can be represented by a utility function. Then agent *k* has other-regarding preferences for the income of agent *j* if his/her utility function, $u^k(y^k, y^j)$ is *not* a constant function of y^j. It is important to distinguish other-regarding preferences from motivations involving trust and reciprocity.

The concepts of trust and positive reciprocity used in this paper are defined as follows. Agent 1 undertakes an action that exhibits "trust" if the chosen action: (a) creates a monetary gain that could be shared with agent 2; and (b) exposes agent 1 to the risk of a loss of utility if agent 2 defects and appropriates too much of the monetary gain. Agent 2 undertakes an action that exhibits "positive reciprocity" if the chosen action: (a) gives agent 1 a monetary gain; and (b) is undertaken instead of an available alternative action that would produce outcomes preferred by agent 2 in the absence of the trusting action by agent 1. Note that the definition of reciprocity incorporates a possible dependence of preferences over outcomes upon the process that generated those outcomes and the perceived intentions of others. This would seem to be a necessary condition for rational agents to undertake actions involving trust and reciprocity. Thus, agent 1 can rationally undertake a trusting action if he/she believes that this choice may trigger a social norm in agent 2 that causes him/her not to defect. Alternatively, agent 1 can rationally undertake a trusting action if he believes that agent 2 has altruistic other-regarding preferences. The experiment design for game triads explained in section 5 makes it possible to discriminate between the implications of other-regarding preferences and trust or reciprocity. This discrimination is central to developing alternatives to the model of self-regarding preferences.

The experimental design explained in section 5 involves game triads that include the investment game introduced by Berg, Dickhaut, and McCabe (1995) and later used by several other authors.

3. The Investment Game

The Berg, Dickhaut, and McCabe (hereafter BDMc) experimental design for the investment game is as follows. Subjects are divided into two groups, the room A group and the room B group. Each individual subject in each group is given ten $1 bills. Each subject in room B is instructed to keep his or her $10. The subjects in room A are informed that each of them, individually, can transfer to an anonymous paired person in room B any number of their own ten

$1 bills, from 0 to all 10, and keep the remainder. Any amount transferred by a room A subject is multiplied by 3 by the experimenter before being delivered to a room B subject. Then each room B subject is given the opportunity to return part, all, or none of the tripled amount of the transfer he or she received from the anonymous paired person in room A.

If one assumes there are no other-regarding preferences, then standard game theory predicts that: (i) room B subjects will keep all of any tripled amounts transferred by room A subjects because room B subjects prefer more money to less; and (ii) knowing this, room A subjects will not transfer any positive amount. This subgame perfect equilibrium of the model of self-regarding preferences is Pareto-inferior to some alternative feasible allocations because it leaves each pair of subjects with $20 when it could have ended up with as much as $40.

Results from investment-game experiments reported by BDMc were that the average amount transferred by room A subjects was $5.16 and the average amount returned by room B subjects was $4.66. When data from this experiment were provided to subjects in a subsequent experiment (the "social history" treatment), the average amount transferred by room A subjects was $5.36 and the average amount returned was $6.46. There was large variability across subjects in the amounts transferred and returned. The experiments reported by BDMc used a "double blind" protocol in which subjects' responses were anonymous to other subjects and the experimenters.

Note what is measured by these experiments. A room A subject may be willing to transfer money to an unknown room B person if he or she *trusts* that some of the tripled amount transferred will be returned. Further, a room B subject may be willing to return part of the tripled amount transferred if he or she practices *positive reciprocity*. But a room A subject may be willing to make a transfer to a paired subject in room B even if there is no opportunity for the latter to return anything. The BDMc design does not provide data that allows one to distinguish between the level of transfers resulting from trust and that resulting from other-regarding preferences. Similarly, their design does not provide data that distinguish between transfers resulting from reciprocity and other-regarding preferences because a room B subject may be willing to transfer money to a paired subject in room A even if that paired subject did not send the money. The experiment design used in the present paper makes it possible to discriminate among transfers motivated by trust, reciprocity, and other-regarding preferences.

4. The Triadic Experimental Design

In order to be able to discriminate empirically among actions motivated by trust, reciprocity, and other-regarding preferences in the investment game, one needs to run experiments with three games. I shall explain the structure of these games and model the players' (utility) payoffs in a general way. Each player's utility function will explicitly incorporate the monetary income of the paired player. It is important to understand that I am *not* assuming that the game players necessarily have other-regarding preferences; instead, I am allowing for that possibility. The subjects' behavior in experiments with the three related games will inform us as to whether they do or do not have other-regarding preferences. The second mover's utility function will explicitly incorporate a variable that introduces the possibility that a trusting action by the paired first mover could trigger an internalized social norm that affects the second mover's utility of the money payoffs from the game. It is also important to understand that I am *not* assuming that the game players necessarily are affected by social norms for reciprocity but am, rather, including that as a possibility. Once again, it is the subjects' behavior in the experiments that will inform us on this question.

4.1 Game A

One of the games, which I shall call game A, is the BDMc investment game that can be modeled as follows. Mover 1 chooses $s^a \in S$, where

(1) $S = \{0,1,2,...10\}$.

The choice of s^a by mover 1 selects the $\Gamma(s^a)$ subgame, in which mover 2 chooses $r^a \in R(s^a)$, where

(2) $$R(s^a) = \{0,1,2,...3s^a\}.$$

At the time mover 1 makes her choice of s^a, she may not know what choice mover 2 will subsequently make. Let the random variable $\tilde{r}$ with probability distribution function $\Omega(\tilde{r} \mid s^a)$, defined on $R(s^a)$, represent mover 1's expectations about the amount of money that will be returned by mover 2 in subgame $\Gamma(s^a)$. Assume that $\Omega(\tilde{r} \mid s^a)$ assigns positive probability to the outcome $r = 0$ for all $s^a \in S$.

Mover 1's expected payoff from choosing s^a in game A is

(3) $$EP_A^1 = \underset{\Omega(\tilde{r}|s^a)}{E}[u^1(10 - s^a + \tilde{r}, 10 + 3s^a - \tilde{r})].$$

In the special case where mover 1 is completely selfish, u^1 is a constant function of the second mover's income.

4.2 Game B

Game B is a dictator game with the same strategy set for the first mover as in the investment game and a specific singleton strategy set for the second mover. Mover 1 chooses $s^b \in S$, where S is defined in statement (1). Mover 2 "chooses" $r^b \in \{0\}$. The payoff to mover 1 is

(4) $$P_B^1 = u^1(10 - s^b, 10 + 3s^b).$$

4.3 Game C(n)

Game C(n) is another dictator game with a specific singleton strategy set for the first mover and the same strategy set for the second mover that he has in the subgame selected in game

A. Mover 1 "chooses" $s^c \in \{s^a\}$. Mover 2 chooses $r^c \in R(s^a)$. The (utility) payoff to mover 2 in game C(n) will not be dependent on the possible operation of a social norm for reciprocity because mover 1 has no decision to make in this game.

4.4 Payoffs Dependent on Social Norms

The utility to mover 2 of the monetary payoffs from a game can be made conditional upon the possible operation of a social norm for reciprocity. Thus, the payoff to mover 2 from the choices of s^a and r^a in game A will be written as

(5) $$P_A^2 = u_{\sigma^a}^2(10 + 3s^a - r^a, 10 - s^a + r^a)$$

because mover 2 knows that mover 1 has chosen the action s^a and may feel obliged to reciprocate. The notation, $u_{\lambda^a}^2$ permits the utility of monetary payoffs to vary with a state variable, λ^a that depends on the amount of money sent by mover 1 to mover 2 in game A:

(6) $$\lambda^a = f(s^a).$$

In contrast, in game C(n) mover 2 knows that mover 1 has no decision to make. Since there is no opportunity for trusting actions by mover 1 in game C(n), there is no reason for a social norm for reciprocating to be triggered. Thus the payoff to mover 2 from the choice of r^c in game C(n) will be written as

(7) $$P_{C(n)}^2 = u^2(10 + 3s^a - r^c, 10 - s^a + r^c).$$

In the special case where a social norm for reciprocity does not affect utility of monetary payoffs, $u_{\lambda^a}^2$ is identical to u^2 for all $s^a \in S$.

4.5 Testing for the Presence of Trust

In order validly to conclude that a first mover has demonstrated trust, the researcher must have knowledge that she has borne a risk of loss from her choice in game A. Thus it must be known that there exists r^z, contained in the support of $\Omega(\tilde{r} \mid s^a)$, and $s^\tau \in S$ such that

(8) $$u^1(10 - s^a + r^z, 10 + 3s^a - r^z) < u^1(10 - s^\tau, 10 + 3s^\tau).$$

Assuming that u^1 is strictly quasi-concave (and recalling that the variables are discrete), the choices by mover 1 allow the researcher to conclude that (8) is satisfied by $r^z = 0$ and $s^\tau = s^b$ if

(9) $$s^a > s^b + 1.$$

This can be seen by noting that the choice by mover 1 in game B and strict quasi-concavity of u^1 imply

(10) $$u^1(10 - s^b, 10 + 3s^b) > u^1(10 - s, 10 + 3s), \forall s \in S, \ni s > s^b + 1.$$

The null hypothesis is that standard game theory makes empirically-correct predictions. In the present context, this means that mover 1 has *not* exhibited trust:

(11) $$H_0^T : s^a \leq s^b + 1,$$

with alternative

(12) $$H_A^T : s^a > s^b + 1.$$

It may seem unlikely that mover 1 will be indifferent between s^b and s^b + 1; hence the null hypothesis in statement (11) may seem to bias the tests against finding that the data contains evidence of trust. Furthermore, across-subjects comparisons between treatments involve means and other aggregations of data for which the $1 unit of discreteness does not apply. Therefore, in the following sections I shall report tests of

(13) $$H_{00}^T : s^a \leq s^b,$$

with alternative

(14) $H_{AA}^{T}: s^{a} > s^{b}$.

Of course, the hypotheses that are tested statistically will be stochastic versions of H_{00}^{T}.

4.6 Testing for the Presence of Reciprocity

In order validly to conclude that a second mover has demonstrated positive reciprocity, the researcher must have knowledge that the second mover has incurred a cost to repay a social debt to the first mover. This can be manifested by the second mover choosing to return an amount of money in game A that: (a) gives the first mover a gain; and (b) in the absence of a social norm for reciprocity, is known to be dispreferred to an available alternative by the second mover. Thus, the second mover has exhibited positive reciprocity in game A if there exists $r^{y} \in R(s^{a})$ such that

(15) $u^{2}(10+3s^{a}-r^{y},10-s^{a}+r^{y}) > u^{2}(10+3s^{a}-r^{a},10-s^{a}+r^{a})$.

Assuming that u^{2} is strictly quasi-concave (and recalling that the variables are discrete), the choices by mover 2 allow the researcher to conclude that (15) is satisfied if

(16) $r^{a} > r^{c}+1$.

This can be seen by noting that the choice by mover 2 in game C(n) and strict quasi-concavity of u^{2} imply

(17) $$\begin{aligned} &u^{2}(10+3s^{a}-r^{c},10-s^{a}+r^{c}) \\ &> u^{2}(10+3s^{a}-r,10-s^{a}+r), \forall r \in R(s^{c}), \ni r > r^{c}+1. \end{aligned}$$

The null hypothesis is that standard game theory makes empirically-correct predictions. In the present context, this means that mover 2 has *not* exhibited reciprocity:

(18) $H_{0}^{R}: r^{a} \leq r^{c}+1$,

with alternative

(19) $H_{A}^{R}: r^{a} > r^{c}+1$.

For the same reasons explained above in the context of testing for trust, the reported tests for reciprocity will be based on stochastic versions of

(20) $H^{R}_{00} : r^{a} \leq r^{c}$,

with alternative

(21) $H^{R}_{AA} : r^{a} > r^{c}$.

5. Experimental Design and Procedures

This experiment includes treatments with games A, B, and C explained in section 4. The instructions for each treatment announce the existence of a second task but do not explain that it is a group decision task involving the investment game.[3] The presence of the second task does not introduce a repeated game because subject anonymity and random matching make it impossible for any subject to acquire a reputation. The experiment sessions are run manually (i.e., not with computers). At the end of a session, a coin is flipped in the presence of the subjects to determine whether task 1 or task 2 has monetary payoff. The payoff procedure is double blind: (a) subject responses are identified only by letters that are private information of the subjects; and (b) monetary payoffs are collected in private from sealed envelopes contained in lettered mailboxes. A detailed explanation of the experiment procedures is contained in appendix 2. An additional appendix, available on request to the author, contains the instructions given to the subjects, the forms used during the experiment, and the questionnaires.

The experiment involves three treatments. Treatment A implements game A, the investment game of BDMc. Each individual in the second-mover group is given 10 one-dollar certificates as a show-up fee. Each individual in the first-mover group is given 10 one-dollar certificates as a show-up fee and given the task of deciding whether he/she wants to transfer to a paired individual in the other group none, some, or all of his/her show-up fee. Any amounts transferred are tripled by the experimenter. Then each individual in the second-mover group is given the task of deciding whether she/he wants to return some, all, or none of the tripled number of certificates received to the paired individual in the other group.

Treatment B implements game B, which involves a decision task that differs from treatment A only in that the individuals in the "second-mover" group do not have a decision to make; thus they do not have an opportunity to return any tokens that they receive. The across-subjects measure of transfers due to trust is then the difference between treatments A and B in the numbers of tokens transferred by first-mover subjects to "second-mover" subjects.

Treatment C implements game C(n), which involves a decision task that differs from treatment A as follows. First, individuals in the "first-mover" group do not have a decision to make. Each individual in the second-mover group is given 10 one-dollar certificates as a show-up fee. Individuals in the "first-mover" group are given show-up fees in amounts equal to the amounts kept (i.e. *not* transferred) by the first-mover subjects in treatment A. The second-mover subjects in treatment C are given "additional certificates" in amounts equal to the amounts received by the second-mover subjects in treatment A from transfers by the first-mover subjects in the latter treatment. The subjects are informed with a table of the exact inverse relation between the number of additional certificates received by a second-mover subject and the show-up fee of the paired "first-mover" subject. The across-subjects measure of transfers due to reciprocity is then the difference between the number of tokens returned in treatments A and C.

The double blind payoffs are implemented by having each subject draw a sealed envelope containing a lettered key from a box containing many envelopes. If the coin flip selects task 1 for monetary payoff, the individuals use their task 1 keys to open lettered mailboxes that contain their monetary payoffs (from task 1 decisions) in sealed envelopes. There is no interaction between experimenters and subjects during decision-making parts of an experiment session. All distribution and collection of boxes containing envelopes for certificates kept and transferred is done by an assistant that is randomly selected from the subject pool. The assistant also announces the outcome of the coin flip and observes all actions of the experimenters in calculating subjects' monetary payoffs.

All of the above design features are common information given to the subjects except for one item. The subjects in treatment C are *not* informed that the amounts of the "first-mover" show-up fees and the second-mover additional certificates are determined by subject decisions in treatment A.

All of the experimental sessions end with each subject being paid an additional $5 for filling out a questionnaire. First-mover subjects and second-mover subjects have distinct questionnaires. The questions asked have three functions: (a) to provide additional data; (b) to provide a check for possible subject confusion about the decision tasks; and (c) to provide checks for possible recording errors by the experimenters and counting errors by the subjects. Subjects did *not* write their names on the questionnaires. The additional data provided by the questionnaires include the subjects' reports of their payoff key letters. Subjects are asked to explain the reasons for their decisions. Data error checks provided by the questionnaires come from asking the subjects to report the numbers of tokens transferred, received, and returned. These reports, together with two distinct records kept by the experimenters, provide accuracy checks on data recording.

6. Comparison with the BDMc Data

One question of interest is whether the first- and second-mover transfers reported by BDMc are robust to the procedures used in the experiments reported here. Data from treatment A will be compared with data from the no-history (NH) and social-history (SH) treatments reported by BDMc.

Treatment NH included 32 subjects run in three experimental sessions. Treatment A included 30 subjects run in two sessions. Both data sets exhibit large variability across subjects. The amount sent varies from 0 to 10 and the amount returned varies from 0 to 20 in the data for both experiments. There are some differences between the two data sets. As reported in the first and second columns of Table 1, on average the subjects in treatment A both sent more ($6.00 vs. $5.16) and returned more ($7.17 vs. $4.66) than the subjects in treatment NH. On average, the sending (or first-mover) subjects in treatment A made a $1.17 profit and those in treatment NH made a $0.50 loss. Also, treatment NH data are noisier than treatment A data in that the former have higher standard deviations for amounts sent and returned and the difference between amounts returned and sent.

The next to last row of Table 1 reports two-sample *t*-tests for differences of means and (non-parametric) Smirnov tests comparing treatment A and treatment NH data. The .840 difference between mean amounts sent is not significantly different from 0 (p = .234) according to the *t*-test. The maximum difference between the cumulative distributions of amounts sent in treatment A and treatment NH is not significant (p > .864) according to the Smirnov test. Since amounts returned are dependent on amounts sent, the next tests are on differences between amounts returned and amounts sent. The 1.67 mean difference between the treatment A and treatment NH return minus sent amounts is not significant (p = .134). Also, the Smirnov test does not find a significant difference between the cumulative distributions of return minus sent amounts in the two treatments (p > .220).

The first and second rows of Table 1 report tobit estimates of the relation between amounts returned and amounts sent in treatment A and treatment NH. The estimated model is given by

$$(22) \qquad R_t = \alpha + \beta S_t + \varepsilon_t ,$$

where R_t is the amount returned by the second mover in subject pair t and S_t is the amount sent by the first mover in pair t. The bounds for the tobit estimation are the bounds imposed by the experiment design:

$$(23) \qquad R_t \in [0, 3S_t].$$

One would expect that the cone created by these bounds might produce heteroskedastic errors. In order to allow for the possibility of heteroskedastic errors, the tobit estimation procedure incorporates estimation of the θ parameter in the following model of multiplicative heteroskedasticity:

(24) $\sigma_t = \sigma e^{\theta S_t}$.

The .055 estimate of the intercept for the treatment A data is not significantly different from 0 (p = .966). The 1.17 slope coefficient is significantly different from 0 (p = .000). Because its standard error is .275, the slope coefficient for the treatment A data is also significantly different from 2 at a 1% significance level. Thus, the subjects' return behavior is significantly different from the prediction of completely-selfish subgame perfect equilibrium ($\beta = 0$) and the equal-split fairness focal point ($\beta = 2$). Sending subjects did, on average, earn a profit on the amounts they sent, but $\hat{\beta}$ is not significantly different from 1, which is the prediction of the zero-loss fairness focal point. The .152 estimate of the parameter, θ of the heteroskedasticity model is significant (p = .004). The right-most column of Table 1 reports the results from a likelihood ratio test for significance of the fitted model. It is highly significant (p < .005) for the treatment A data.

The .667 estimate of the intercept for treatment NH data is not significantly different from 0 (p = .692). The .687 slope coefficient for treatment NH data is significantly different from 0 (p = .082). Because its standard error is .394, the slope coefficient is also significantly different from 2 at a 1% significance level. Thus, the subjects' return behavior is significantly different from the prediction of completely-selfish subgame perfect equilibrium ($\beta = 0$) and the equal-split fairness focal point ($\beta = 2$). Sending subjects did, on average, make a loss on the amounts they sent, but $\hat{\beta}$ is not significantly different from 1. The .140 estimate of the parameter of the heteroskedasticity model is not significant (p = .175). The result of the likelihood ratio test is insignificance (p > .100).

The next to last row of Table 1 reports tobit estimates of the model,

(25) $$R_t = \alpha + \beta S_t + \gamma D_t S_t + \varepsilon_t$$

where:

(26) D_t = 1 for treatment A data

= 0 for treatment NH data.

This estimation uses the bounds and heteroskedasticticity model given by equations (23) and (24). The estimate of γ is not significantly different from 0 (p = .179), which provides additional support for the conclusion that the differences between the data from treatment NH and treatment A are not statistically significant.

6.2 Treatment A vs. Social-History BDMc Data

Treatment SH included 28 pairs of subjects run in three experimental sessions. Treatment SH data looks even more similar to treatment A data than does treatment NH data, thus it is not surprising that no significant differences between the treatment A and treatment SH data are detected by the various tests reported in Table 1.

6.3 Conclusions Supported by Data from All Three Treatments

The finding reported here of no significant differences between data from treatment A and data from the two BDMc experiments provides further support for the conclusion by Ortmann, Fitzgerald, and Boeing (2000) that behavior in the investment game is robust to alternative experimental designs. Treatment A data, treatment NH data, and treatment SH data all support these three conclusions: (a) the data are inconsistent with the subgame perfect equilibrium of the model of self-regarding preferences; (b) the data are inconsistent with the equal-split fairness focal point; and (c) the data are *not* significantly different from the zero-loss fairness focal point.

7. Trust and Other-Regarding Preferences

Data from treatments A and B can be used to decompose the total amounts sent by first movers into amounts sent because of trust and amounts sent because of other-regarding preferences. The tests reported in Table 2 use data for all except four of the subjects in sessions with treatment B. Tests with other subject samples, and the reasons for deleting data for four subjects in the tests reported in Table 2, are reported in appendix 1.

The first row of Table 2 reports the mean amounts sent by first movers in treatments A and B. The mean amount sent was slightly larger in treatment A than in treatment B but the difference is insignificant ($p = .389$). The result from the Smirnov test is similar to that from the *t*-test, with no significant difference between treatments A and B ($p > .10$). Thus there is no evidence in these data that the first-mover subjects sent part of their show-up fees to the paired second-mover subjects *because of* a trust that the second movers would not defect.

8. Reciprocity and Other-Regarding Preferences

Data from implementations of games A and C can be used to discriminate between amounts returned by second movers due to other-regarding preferences and amounts returned because of positive reciprocity.

The first row of Table 3 reports the mean differences for all subjects between amounts returned and amounts "sent" in treatments A and C. On average, the amount returned in treatment A exceeded the amount sent by $1.17. Thus the second movers did share the profit provided by the first movers' transfers. The average amount returned in treatment C was $1.23 less than the amount "sent." The difference between the treatment A and treatment C outcomes is in the direction implied by positive reciprocity. The one-tailed *t*-test for difference in means implies that the difference between the return minus sent means for treatments A and C is significantly greater than 0 (p = .017). In contrast, the Smirnov test does not find a significant difference between the cumulative distributions of return minus sent amounts in treatments A and C.

The sixth through tenth columns of the first row of Table 3 report tobit estimates using data for all subjects of the parameters of the following relation between amounts sent and amounts returned in treatments A and C:

(27) $$R_t = \alpha + \beta D_t S_t + \gamma S_t + \varepsilon_t ,$$

where

(28) D_t = 1 for treatment A data

= 0 for treatment C data.

This estimation uses the bounds and heteroskedasticity model given by equations (23) and (24). Note that $\hat{\beta}$ is the estimate of the effect of reciprocity on amount returned by second movers to first movers. We observe that $\hat{\beta}$ is positive and significantly greater than 0 (p = .002). Thus the data provide support for behavior involving positive reciprocity. Using the mean amounts returned in the investment and dictator games, one concludes that 67% of the total amount returned by second movers to first movers in the investment game is motivated by other-regarding preferences and 33% is motivated by positive reciprocity.

Berg, Dickhaut, and McCabe (1995) noted that first movers who send all \$10, and perhaps also those who send \$5, might elicit larger returns by second movers. This possibility was examined for the data from the experiment reported here by introducing dummy variables for \$5 and \$10 amounts sent into the tobit model. Neither of the estimated coefficients for the \$5 and \$10 dummy variables is significantly different from zero.

9. Concluding Remarks

This paper reports experiments with game triads that include the investment game. Several researchers had previously established the replicable result that the majority of first movers send positive amounts and the majority of second movers return positive amounts in investment game experiments. This pattern of results, and results from many other non-market fairness experiments, are inconsistent with the special case of game theory in which players' utilities are increasing functions of only their own monetary payoffs in an experiment. This leaves the profession with the task of constructing a less restrictive theory of utility that can maintain consistency with the empirical evidence. But this task cannot be undertaken successfully unless we can discriminate among the observable implications of alternative possible motivations. The game triad experiments reported here make it possible to discriminate among the observable implications for subjects' choices of trust, reciprocity, and other-regarding preferences in the investment game.

The experiment does not provide any evidence of trusting behavior. The experiment does provide some evidence of reciprocity but, nevertheless, 67% of the amount returned by second movers to first movers can be attributed to other-regarding preferences.

Results from the triadic design for experiments with the investment game demonstrate the importance of conducting experiments with designs that can discriminate among the motivations that are alternatives to that of wanting only to maximize one's own money income. Play in this game is characterized by both other-regarding preferences and reciprocity. This observation is central to the implication that data from the investment game have for guiding the development of utility theory to increase the empirical validity of game theory. The implication is that the data demand the inclusion of both other-regarding preferences for outcomes and beliefs about others' intentions. There are a few other studies that have used control treatments for intentions.

Blount (1995) compared second mover rejections in a standard ultimatum game with second mover rejections in games in which the first move was selected randomly or by an outside party rather than by the subject that would receive the first mover's monetary payoff. She found lower rejection rates in the random treatment than in the standard ultimatum game but no significant difference between the rejection rates in the third party and standard games. Charness (1996) used Blount's control treatments in experiments with the gift exchange game. He found somewhat *higher* second mover contributions in the outside party and random treatments than in the standard gift exchange game. Bolton, Brandts, and Ockenfels (1998) experimented with an intentions-control treatment in the context of simple dilemma games. In the control treatment, the row player "chooses" between two identical rows of monetary payoffs. They found no significant differences between the column players' responses in the control treatment and the positive and negative reciprocity treatments.

The general conclusion from the research reported in the present paper is about experimental methods: We cannot learn what we need to know in order to provide empirical guidance to theory development by designing single-game experiments to test theoretical predictions derived from the assumption that subjects' utilities are completely selfish. We need to use designs that, like the triadic design, can discriminate between various motivations so that we can discover the various contexts in which trust, reciprocity, and/or other-regarding preferences are significant determinants of behavior and the contexts in which models of traditional economic man will suffice to predict behavior.

Endnotes

* Financial support was provided by the Decision Risk and Management Science Program, National Science Foundation (grant number SES-9818561). I am grateful for helpful comments and suggestions to the students in my Ph.D. seminar on experimental economics, most especially Cary A. Deck and Mary L. Rigdon, and to Professors Ronald L. Oaxaca, Vernon L. Smith, and John Wooders. Valuable research assistance was provided by Mary L. Rigdon, Marisa D. Bernal, Matthew E. Cox, Cary A. Deck, and Supriya Sarnikar.

1. The research program includes the following papers: Berg, Dickhaut, and McCabe (1995); Hoffman, McCabe, Shachat, and Smith (1994); Hoffman, McCabe, and Smith (1996, 1998); McCabe, Rassenti, and Smith (1996, 1998); and Smith (1998).

2. The research program includes the following papers: Fehr and Falk (1999), Fehr and Gächter (2000a,b), Fehr, Gächter, and Kirchsteiger (1996, 1997), and Fehr, Kirchsteiger, and Riedl, (1993).

3. Putterman and Ben-Ner (1998) have previously used the labels “self-regarding” and “other-regarding” to describe a similar distinction between narrowly selfish preferences and preferences that can be altruistic.

4. The group data are reported in Cox (2000).

References

Berg, Joyce, John Dickhaut, and Kevin McCabe, "Trust, Reciprocity, and Social History." *Games and Economic Behavior*, July 1995, 10(1), pp. 122-42.

Blount, Sally, "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences." *Organizational Behavior and Human Decision Processes*, August 1995, 63(2), pp. 131-44.

Bolton, Gary E., Jordi Brandts, and Axel Ockenfels, "Measuring Motivations for the Reciprocal Responses Observed in Simple Dilemma Games." *Experimental Economics*, 1998, 1(3), pp. 207-19.

Bolton, Gary E. and Axel Ockenfels, "ERC: A Theory of Equity, Reciprocity and Competition." *American Economic Review*, March 2000, 90(1), pp. 166-93.

Charness, Gary, "Attribution and Reciprocity in a Simulated Labor Market: An Experimental Study." Working Paper, University of California at Berkeley, 1996.

Cox, James C., "Trust, Reciprocity, and Other-Regarding Preferences: Groups vs. Individuals and Males vs. Females." University of Arizona discussion paper, September 1999.

Dufwenberg, Martin and Georg Kirchsteiger, "A Theory of Sequential Reciprocity." discussion paper, CentER for Economics Research, Tilburg University, 1999.

Falk, Armin and Urs Fischbacher, "A Theory of Reciprocity." Working Paper No. 6, Institute for Empirical Research in Economics, University of Zurich, 1999.

Fehr, Ernst and Armin Falk, "Wage Rigidity in a Competitive Incomplete Contract Market." *Journal of Political Economy*, 107, 1999, pp. 106-134.

Fehr, Ernst and Simon Gächter, "Do Incentive Contracts Crowd Out Voluntary Cooperation?" University of Zurich, February 2000a.

Fehr, Ernst and Simon Gächter, "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives*, Summer 2000b, 14(3), pp. 159-81.

Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger, "Reciprocal Fairness and Noncompensating Wage Differentials." *Journal of Institutional and Theoretical Economics*, 152, 1996, pp. 608-640.

Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger, "Reciprocity as a Contract Enforcement Device: Experimental Evidence." *Econometrica*, July 1997, 65(4), pp. 833-60.

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl, "Does Fairness Prevent Market Clearing? An Experimental Investigation." *Quarterly Journal of Economics*, May 1993, 108(2), pp. 437-60.

Fehr, Ernst and Klaus M. Schmidt, "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, August 1999, 114(3), pp. 817-68.

Hoffman, Elizabeth, Kevin A. McCabe, Keith Shachat, and Vernon L. Smith, "Preferences, Property Rights, and Anonymity in Bargaining Games." *Games and Economic Behavior*, November 1994, 7(3), pp. 346-80.

Hoffman, Elizabeth, Kevin A. McCabe, and Vernon L. Smith, "Social Distance and Other-Regarding Behavior in Dictator Games." *American Economic Review*, 3, 1996, pp. 653-660.

Hoffman, Elizabeth, Kevin A. McCabe, and Vernon L. Smith, "Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology." *Economic Inquiry*, 36, 1998, pp. 335-352.

Levine, David K., "Modeling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics*, July 1998, 1(3), pp. 593-622.

McCabe, Kevin A., Stephen J. Rassenti, and Vernon L. Smith, "Game Theory and Reciprocity in Some Extensive Form Bargaining Games." *Proceedings of the National Academy of Sciences*, 93, November 1996, pp. 13421-13428.

McCabe, Kevin A., Stephen J. Rassenti, and Vernon L. Smith, "Reciprocity, Trust, and Payoff Privacy in Extensive Form Bargaining Games." *Games and Economic Behavior*, 24, 1998, pp. 10-24.

Nowak, M.A. and K. Sigmund, "Evolution of Indirect Reciprocity by Image Scoring." *Nature*, 393, pp. 573-7.

Nowak, M.A. and K. Sigmund, "The Dynamics of Indirect Reciprocity." *Journal of Theoretical Biology*, 194, pp. 561-74.

Ortmann, Andreas, John Fitzgerald, and Carl Boeing, "Trust, Reciprocity, and Social History: A Re-examination." *Experimental Economics*, 3, 2000, pp. 81-100.

Putterman, Louis and Avner Ben-Ner, (eds.), *Economics, Values, and Organization*, New York, NY: Cambridge University Press.

Rabin, Matthew, "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83, 1993, pp. 1281-1302.

Smith, Vernon L., "Distinguished Guest Lecture: The Two Faces of Adam Smith." *Southern Economic Journal*, 65, 1998, pp. 1-19.

von Neumann, John and Oskar Morgenstern, first edition (1944), second edition (1947). *Theory of Games and Economic Behavior*, Princeton, NJ: Princeton University Press.

Table 1. Comparison of Treatment A and BDMc Treatments NH and SH

Data	*Sent Mean*	*Returned Mean*	Ret. Sent Mean	Sent Means Test	Ret. – Sent Means Test	Sent Smirnov Test	Ret. – Sent Smirnov Test	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$	LR Test
Tr. A	6.00 [2.59]	7.17 [4.82]	1.17 [3.71]	…..	…..	…..	…..	.055 (.966)	1.17 (.000)	…..	.152 (.004)	25.1 (.000)
Tr. NH	5.16 [2.94]	4.66 [5.55]	-.50 [4.89]	…..	…..	…..	…..	.667 (.692)	.687 (.082)	…..	.140 (.175)	3.56 (> .100)
Tr. SH	5.36 [3.53]	6.46 [6.19]	1.11 [4.31]					.275 (.858)	1.11 (.001)	…..	.096 (.160)	27.1 (.000)
Tr. A vs. Tr. NH	…..	…..	…..	.840 (.234)	1.67 (.134)	.574 (>.864)	1.01 (>.220)	.479 (.650)	.747 (.003)	.345 (.179)	.136 (.007)	23.0 (< .005)
Tr. A vs. Tr. SH	…..	…..	…..	.795 (.430)	.057 (.955)	.833 (>.465)	.520 (>.923)	.163 (.861)	1.13 (.000)	.026 (.897)	.124 (.003)	53.6 (.000)

p-values in parentheses.
Standard deviations in brackets.

Table 2. Decomposition Tests

Data	Sent Mean	Returned Mean	Means Tests	Smirnov Tests
Tr. A	6.00 [2.59] {30}	7.17 [4.82] {30}	…	…
Tr. A vs. Tr. B	…	…	.184 (.389) [1]	.100 (> .10) [1]
Tr. B	5.81 [2.73] {38}	…	…	…
Tr. A vs. Tr.C Ret. - Sent		…	2.07 (.017) [1]	.133 (>.10) [1]
Tr. C	…	5.10 [5.27] {30}	…	…

p-values in parentheses.
[1] denotes a one-tailed test.
Standard deviations in brackets.
Number of subjects in braces.

Table 3. Tobit Analysis of Return Behavior

$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$	LR Test
-.002 (.998)	.427 (.002)[1]	.748 (.000)	.181 (.000)	23.0 (< .005)

p-values in parentheses.
[1] denotes a one-tailed test.

Appendix 1. Analysis of Data for Alternative Subject Samples

During the first session involving treatment B, the questionnaire responses of three subjects indicated that they did not fully understand the decision task. This caused the experimenter to revise the treatment B subject information form before running subsequent sessions with this treatment.

In the last session with treatment B, there was one repeat subject. The experimenter recognized the subject as a repeat subject shortly after the session began. The double blind experimental protocol was discreetly compromised in order to learn this subject's identification letter (but not her name). This was done by silently counting the number of questionnaires that were sequentially deposited by the subjects in a stack in a collection box. It was determined that this subject's questionnaire was in position *n* in the stack. The subjects were not aware this was being done and all of their decisions had already been completed at the time the counting was done. This was the last session in this treatment.

Table 2 uses data for all subjects except the three possibly-confused subjects and one repeat subject. An argument can be made that data for all subjects should be used. An argument can also be made that data for all of the subjects in the first session with treatment B (using the original subject information form) should be deleted. The first row of Table 2A reports analysis of data for all subjects. The second row reports results from analysis of the data that remains after excluding all subjects from the first session with treatment B and the repeat subject. The conclusions are the same as from the analysis reported in Table 2.

Table 2A. Money Sent Due to Trust and Other-Regarding Preferences

Subject Sample	Treatment A Mean	Treatment B Mean	Means Test	Smirnov Test
All	6.00 [2.59] {30}	6.14 [2.82] {42}	-.143 (.588)	.100 (> .10) [I]
Excl. First Session and Repeat	6.00 [2.59] {30}	6.23 [2.72] {26}	-.231 (.626) [I]	.167 (> .10) [I]

p-values in parentheses.
[I] denotes a one-tailed test.
Standard deviations in brackets.
Number of subjects in braces.

Appendix 2. Experiment Procedures

The subjects first assembled in the sign-in room of the Economic Science Laboratory and recorded their names, student identification numbers, and signatures on a form. Then a monitor was chosen randomly from the subject sample (by drawing a ball from a bingo cage) and given the responsibility of ensuring that the experimenters followed the procedures contained in the subject instructions for calculating money payoffs. The monitor was paid $20 for this job. The other subjects were not informed of the amount of this payment in order to avoid the possible creation of a focal earnings figure. Next the subjects were randomly divided into two equal-size groups, called group X and group Y and escorted into the large room of the Economic Science Laboratory. The procedures differed somewhat across the three treatments because of the properties of the experiment design

Treatment A involves the BDMc investment game modeled in subsection 4.1 above. In a treatment A session, the group X subjects were seated at widely-separated computer terminals with privacy side and front partitions. (The computers were not used.) The group Y subjects were standing at the back of the room at the beginning of the session with treatment A. Each group Y subject was given an envelope labeled “my show-up fee” that contained ten task one $1 certificates. Each subject and the monitor were given copies of the instructions for “task one” (the individual decision task). Then an experimenter read aloud the instructions. The instructions for all treatments are contained in an appendix available upon request. After the reading of the instructions was completed, the group Y subjects were escorted back to the sign-in room by one of the experimenters. (The group X subjects had no further contact with the group Y subjects until after all decisions in both decision tasks had been completed.) Then the group X subjects were given the opportunity to raise their hands if they had questions. If a subject raised his hand, he was approached by an experimenter and given an opportunity to ask questions and receive answers in a low voice that could not be overheard by other subjects. When there were no more questions, the experimenter left the room and the monitor took over to conduct the first-mover individual decision task with the group X subjects.

The monitor carried a large box that contained smaller boxes equal in number to the number of subjects. Each subject was given the opportunity to point to any remaining small box to indicate she wanted that one. (The boxes all looked the same to the experimenters.) A subject opened her box to find an envelope labeled "my show-up fee" that contained ten task one $1 certificates. The box also contained an empty envelope labeled "certificates sent to a paired person in group Y" and an envelope containing a lettered task one mailbox key. Finally, the box contained a one-page form that summarized the nature of the first-mover individual decision task. This form and the corresponding forms for other treatments are contained in an appendix available upon request. All envelopes in the box were labeled with the letter on the mailbox key.

Subjects were given 10 minutes to complete this task. When a subject was finished, he put all of the envelopes except the key envelope back in the box and summoned the monitor to collect the box. The monitor then carried the large box full of small boxes into another room for data recording and the preparation of boxes for the group Y, second-mover subjects. The monitor witnessed all data recording and group Y box preparations.

While the boxes were being processed, one experimenter escorted the group X subjects out a side door of the Economic Science Laboratory and down the hall to the breakout rooms of the Decision Behavior Laboratory. Next, another experimenter escorted the group Y subjects into the Economic Science Laboratory to get ready for their second-mover decisions in the individual decision task.

The group Y subjects were given boxes by the monitor. Each box contained an envelope with a lettered task one mailbox key. The box contained two empty envelopes, one labeled "my certificates" and the other labeled "certificates returned to the paired person in group X." The box contained the tripled number of certificates sent by the paired person in group X and a form summarizing the decision task. The form is contained in an appendix available upon request. The group Y subjects had to decide how many of the certificates to put in the envelopes labeled "my certificates" and "certificates returned to the paired person in group X." The group Y subjects were given 10 minutes to complete the task. When a subject was finished, she put all envelopes except the key envelope back in the box and summoned the monitor to collect it. The monitor then carried the large box of little boxes to another room and watched the data recording

The second-mover decisions in task one were conducted simultaneously with the first-mover decisions in task two. The first-mover decisions in task two were made by three-person committees that were formed by the experimenter by the order in which the subjects entered the laboratory from the hallway. Thus, the first three subjects were assigned to be in the first committee, the next three in the second committee, and so on. Each committee was seated in its own small breakout room. Each member of each committee was given the written subject instructions for task two. Then an experimenter read aloud the instructions while all breakout room doors remained open. Subjects were then given the opportunity to indicate whether they had any questions. If there was a question, the experimenter entered the appropriate breakout room and closed the door before the question was asked and answered. When there were no more questions, the experimenter left and the monitor took over. The monitor permitted the members of each committee to point to a small box contained in a large box to indicate which remaining box the committee wanted. A committee's box contained an envelope labeled "our show-up fee" that contained 30 task two $1 certificates. The box also contained an envelope labeled "certificates sent to a paired committee in group Y" and an envelope containing a lettered task two mailbox key. Finally, the box contained a one-page summary of the group decision task. The form is contained in an appendix available upon request. The committees were given 20 minutes to complete their tasks. When a committee was finished, it put all envelopes except the key envelope back in the box and summoned the monitor by opening the door to its breakout room. The monitor carried the large box full of little boxes to the processing room and watched the data recording and preparation of boxes for the group Y committees. Next, an experimenter escorted the group X subjects back to the sign-in room. After all of the group X subjects were in the sign-in room and the door was closed, an experimenter escorted the group Y subjects out a side door of the Economic Science Laboratory and down the hallway to the breakout rooms of the Decision Behavior Laboratory.

The group Y subjects then made their task two, second-mover decisions. Each group Y committee was given an envelope labeled "our show-up fee" that contained 30 task two $1 certificates. The procedures for reading instructions, answering questions, and the role of the monitor were like those for the first-mover, group X subjects. Each group Y committee's box contained the tripled number of certificates sent to it by the paired committee in group X. The box also contained an envelope with a task two key, a summary instruction form, and two empty envelopes. The empty envelopes were labeled "our certificates" and "certificates returned to the paired committee in group X."

After the group Y committees finished their task two decisions, they were escorted back down the hall to rejoin the group X subjects in the Economic Science Laboratory. Next, an experimenter flipped a coin in the presence of all of the subjects and the monitor. The monitor announced whether the coin came up heads or tails. If heads (tails) then each task one (Two) $1 certificate was exchanged for one United States dollar. While the subjects' money payoffs were calculated, they filled out the questionnaires. In addition to the salient money payoff, each subject was paid $5 upon depositing her completed questionnaire in a box. After the questionnaires were completed the group X subjects went together to obtain sealed envelopes containing their money payoffs from lettered mailboxes. They had been asked to exit the building after obtaining their envelopes and not to open their envelopes until out of the building. After the group X subjects had left, the group Y subjects obtained their payoff envelopes from the lettered mailboxes.

The procedures for treatment B differed as follows from the treatment A procedures explained above. The group Y subjects did not make a decision in task one. The procedures for treatment C differed as follows from those for treatment A. At the beginning of task one, the group Y subjects were seated in the Economic Science Laboratory and the group X subjects were standing at the back. The group X subjects did not make a decision in task one.